

# Accounting for Misclassification in Multispecies Distribution Models.

Kwaku Peprah Adjei <sup>\*1,2</sup>, Robert Bob O'Hara <sup>†1,2</sup>, Anders Finstad <sup>‡2</sup>, and Wouter Koch <sup>§2</sup>

<sup>1</sup>Department of Mathematical Sciences, Norwegian University of Science and Technology NTNU,  
7491 Trondheim, Norway

<sup>2</sup>Centre for Biodiversity Dynamics, Norwegian University of Science and Technology NTNU,  
7491 Trondheim, Norway

## Abstract

- Species identification errors may have severe implications for the inference of species distributions. Accounting for misclassification in species distributions is an important topic of biodiversity research. With an increasing amount of biodiversity that comes from Citizen Science projects, where identification is not verified by preserved specimens, this issue is becoming more important. This has often been dealt with by accounting for false positives in species distribution models. However, the problem should account for misclassifications in general.
- Here we present a flexible framework that accounts for misclassification in the distribution models and provides estimates of uncertainty around these estimates. The model was applied to data on viceroy, queen and monarch butterflies in the United States. The data were obtained from the iNaturalist database in the period 2019 to 2020.
- Simulations and analysis of butterfly data showed that the proposed model was able to correct the reported abundance distribution for misclassification and also predict the true state for misclassified state.

**Key words:** Misclassification, Bayesian models, MCMC, Multispecies, False positives, Citizen Science

---

\*kwaku.p.adjei@ntnu.no

†bob.ohara@ntnu.no

‡anders.finstad@ntnu.no

§wouter.koch@artsdatabanken.no

# 1 Introduction

Species observations obtained from Citizen Science programs where non-professionals are involved in data collection and other scientific procedures is rapidly becoming an essential part of the knowledge base for ecological conservation. Data collected by these hobbyists has led to a massive increase in the coverage and amount of data collected. However, misclassification of this data remains a critical issue to consider when parameterizing species distribution models (hereafter noted as SDMs) based in part or fully on Citizen Science (hereafter noted as CS) data [Bird et al., 2014].

The problem of misclassification has been construed as problems of false positives and false negatives, which are two common sources of biases in biodiversity data. False positives occur when non-targeted species are reported as the species of interest [Clare et al., 2021]; and false negatives occur when targeted species are reported as non-targeted species or a species is not observed. Many approaches have been developed to account for these false positives and negatives in species distribution models [Kéry and Royle, 2020, Wright et al., 2020, Chambert et al., 2015, Royle and Link, 2006]. Out of these, very few are from multi-species distribution models [Wright et al., 2020, Chambert et al., 2018a]. However, there has been increasing literature that supports the assertion that accounting for false positives in single species models are not adequate. This holds especially in instances where the target species can be easily classified as some other species or co-occur with other species [Wright et al., 2020]. Failure to account for the false positives lead to underestimation of species ranges and biases in estimates of species occurrence or abundance [Molinari-Jobin et al., 2012].

Here we present a method to deal with misidentifications that, rather than classifying an observation as true or false positive, tries to classify an observation as the correct species. The method we present needs data with both an observation and validation. These validation data can be from Machine learning algorithms such as Artsorakel [Centre, 2022] or by human experts such as community verification for iNaturalist data [Matheson, 2014]. Although the developments in ML has led to improvements in classification of individuals reported [Willi et al., 2019, Keshavan et al., 2019, Saoud et al., 2020], they do not usually provide uncertainty in the estimates of the classification probabilities. Here we view the classifications (either by ML or community verification) as a verified data and use our method to account for the misclassification. The additional verified data makes the estimation of the misclassification probabilities identifiable like any other occupancy or abundance model [Wright et al., 2020, Chambert et al., 2015].

There are instances where observed species are recorded at a higher taxonomic level such as genus level or a different level (for example "unidentified woodpecker"). Such possibilities have not been considered in the existing model frameworks that account for misclassification in SDMs. We therefore propose a flexible and general framework that allows inter-taxonomic (mis)classification such as species - genus (mis)classification to be accounted for.

Another pending question is whether the probability of correct identification can be predicted for already reported data. Our proposed model framework makes it possible to continually update the information of the verified species

by taking into account the estimated classification probabilities from available verified identifications. This can help in the identification of individuals, for example, if an individual is misidentified as being well out of its range.

The aim of the present is to propose a multi-species distribution model in two parts. First, handling the misclassification with uncertainties around them. Secondly, another part that predicts the true identity of the observed data. This model we propose is very flexible and general in its application.

## 2 Methodology

### 2.1 Model framework

The proposed model framework assumes that each observation is of an individual with a true state (usually unknown) and observed state ( $\mathbf{Y}$ ) (where the sets of possible states do not have to be the same), and for that observation there is a verified state ( $\mathbf{V}$ ) which we assume to be equal to the true state.

We assume that over a given study region  $D \in \mathbb{R}^2$ , there are  $i = 1, 2, \dots, R$  indexed locations, where each location has only one of the  $s = 1, 2, \dots, S$  true individuals. We also assume that these locations are the realisations of the point pattern which is characterised by the mean intensity [Cressie, 2015]. Here we model the mean intensity  $\lambda_{is}$  of the true state  $s$  as an inhomogeneous point process (PPM) which assumes that the data are dependent on the environment covariate, or as a log-Cox Gaussian Point process (LGCP) where we assume a spatial dependency in the data [Renner et al., 2015]. The mean intensity of the verified observations ( $\mathbf{V}$ ) is modelled using the inhomogeneous point process and defined as:

$$\ln(\lambda_{is}) = \beta_{0s} + \beta'_s \mathbf{x}, \quad (1)$$

where  $\beta_{0s}$  is the intercept of verified state  $s$ ,  $\beta_s$  is the vector of covariate effect and  $\mathbf{x}$  is the covariate that affects the observation of the verified (true) states.

Let  $p_{is}$  be the probability of obtaining verified state  $s$  at location  $i$ . Estimating these probabilities from the mean intensities of the states,

$$p_{is} = \frac{\lambda_{is}}{\sum_s \lambda_{is}}, \quad (2)$$

where  $\lambda_{is}$  is given by (1).

The verified states  $s = 1, 2, \dots, S$  at location  $i = 1, 2, \dots, R$  is then a realisation from a categorical distribution with probability  $p_{is}$ . This particular specification of the model for the true states is the loglinear model and can be easily adjusted into multinomial logit model used for categorical variables (See Supplementary information 1).

Assuming we have verified data which represent the true states, each reported observation can then be seen as a draw from any of  $K$  categories under consideration with a given probability. These categories could be on any taxonomic level, from the species level to the kingdom level, partial identifications or unidentified groups. For example, a set of hypothetical ornithological reportings could have four verified states: black woodpecker, pileated woodpecker, yellow-bellied sapsucker and Williamson’s sapsucker. As well as being reported at the species level, more general classifications can be used. For example, the North American Breeding Bird Survey also includes ”unidentified Sapsucker”, ”unidentified Woodpecker”, and four different Northern Flickers [Sauer et al., 2013]. An example of the classification probability (confusion matrix) is shown in Table 1.

Verified state	Observed state			
	black wood-pecker	woodpecker	sapsucker	Unidentified woodpecker
black woodpecker	0.8	0.1	0.1	0
pileated woodpecker	0.05	0.7	0.2	0.05
yellow-bellied sapsucker	0	0.9	0	0.1
williamson’s sapsucker	0	0	1	0

Table 1: Example of confusion matrix that is applicable to our model.

Let  $\Omega_{sk}$  be the probability that a single verified state  $s \in \{1, \dots, S\}$  is classified as category  $k \in \{1, \dots, K\}$ . The probabilities across all the possible  $k$  categories sum to 1. Generally, the confusion matrix for the classification can be expressed as:

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{11} & \cdots & \Omega_{1k} & \cdots & \Omega_{1K} \\ \Omega_{21} & \Omega_{22} & \cdots & \Omega_{2k} & \cdots & \Omega_{2K} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \Omega_{s1} & \Omega_{s2} & \cdots & \Omega_{sk} & \cdots & \Omega_{sK} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \Omega_{S1} & \Omega_{S2} & \cdots & \Omega_{Sk} & \cdots & \Omega_{SK} \end{bmatrix} \quad (3)$$

where the rows correspond to the verified states  $s$  and the columns correspond to the reported categories  $k$ .

Given which verified state was present, the reported category at location  $i$ ,  $Y_i$ , is a draw from  $K$  categories with probability  $\Omega$ .

$$Y_i|V_i \sim \text{Categorical}(\Omega_{\cdot,k}), \quad (4)$$

where  $\Omega_{\cdot,k}$  refers to the probability of being classified as any of the  $K$  categories given which verified state was present at location  $i$ .

The hierarchical likelihood of the model proposed is given as follows:

$$\begin{aligned}\ln(\lambda_{is}) &= \mathbf{x}'_i \beta \\ p_{is} &= \frac{\lambda_{is}}{\sum_i \lambda_{is}} \\ V_i | p_{is} &\sim \text{Categorical}(p_{i.}) \\ Y_i | V_i &\sim \text{Categorical}(\Omega_{V_i, .})\end{aligned}$$

for  $s = 1, \dots, S$  verified states,  $k = 1, 2, \dots, K$  reported categories and  $i = 1, 2, \dots, R$  number of sites.

## 2.2 Generalisation of model framework

The proposed framework generalises the existing multi-species distribution models that account for false positives. Given our specification of model framework, our model is connected to that of Wright et al. [2020] by using the relationship between the multinomial and the Poisson distribution [Steel et al., 1953] as well as using a species by species confusion matrix (Supplementary information 2). By extending the model proposed by Wright et al. [2020], our model framework can also be seen as a generalisation of the works by Chambert et al. [2018a], Chambert et al. [2015] and Chambert et al. [2018b].

## 2.3 Predicting true states abundance

The true states can also be predicted from the estimates of  $\Omega$  to check whether we obtain each true state intensity and abundance map and to explore how the relationship between the observed and true states changes with the environmental changes (which only affects the true distributions). With this result, species abundance maps for regions yet to be covered could be predicted from available CS data and the classification probabilities.

Let  $\Gamma_{iks}$  be the probability that the true (verified) state is  $s$  given that category  $k$  was observed at site  $i$ . From Bayes theorem,

$$\Gamma_{iks} = \frac{P(\text{reported} = k | \text{verified} = s) P(\text{verified} = s)}{\sum_s P(\text{reported} = k | \text{verified} = s) P(\text{verified} = s)} = \frac{\hat{\Omega}_{sk} \times \frac{\hat{\lambda}_{is}}{\sum_i \lambda_{is}}}{\sum_s \left[ \hat{\Omega}_{sk} \times \frac{\hat{\lambda}_{is}}{\sum_i \lambda_{is}} \right]}, \quad (5)$$

where  $\hat{\Omega}_{sk}$  is the estimated probability of reporting category  $k$  given the verified state  $s$  and  $\hat{\lambda}_{is}$  is the estimate mean intensity of the verified state for  $s = 1, 2, \dots, S$  verified species,  $k = 1, 2, \dots, K$  reported category and  $i = 1, 2, \dots, R$  indexed sites.

The predicted true states can be chosen as the species with the maximum of the probability  $\Gamma_{iks}$  for all verified states  $i = 1, 2, \dots, S$ . Alternatively, the true states can be simulated from Categorical distribution with probability

$(\Gamma_{k,j})$  for the reported categories  $k = 1, 2, \dots, K$  at location  $j$ . Thus, the predicted true state depends on the estimates of  $\Omega$  and the predicted mean intensity of each state at the location  $j$ .

## 2.4 Simulation Study

To demonstrate how our model works and its use in prediction, we performed a simulation study for  $S = 3$  verified states and  $K = 4$  reported categories over  $R = 400$  indexed sites. A single covariate was used and was generated uniformly over the domain  $[-1, 1]$ . The true intensity was simulated from equation (1). The intercept of the model for the three verified states were chosen as  $\beta_{01} = -1$ ,  $\beta_{02} = 1$ , and  $\beta_{03} = 0$ . The environmental covariate effect for each verified state was chosen as  $\beta_{11} = 2$ ,  $\beta_{12} = 0$  and  $\beta_{13} = -2$ . The intercept and covariate effect for the three verified states were carefully chosen such that they have a sum to zero constraint, and all species would have a location where they are more probable to occur. The range of covariate values with the probability of observing a verified (true) state  $i$  from our parameter values chosen is shown in Figure 1.

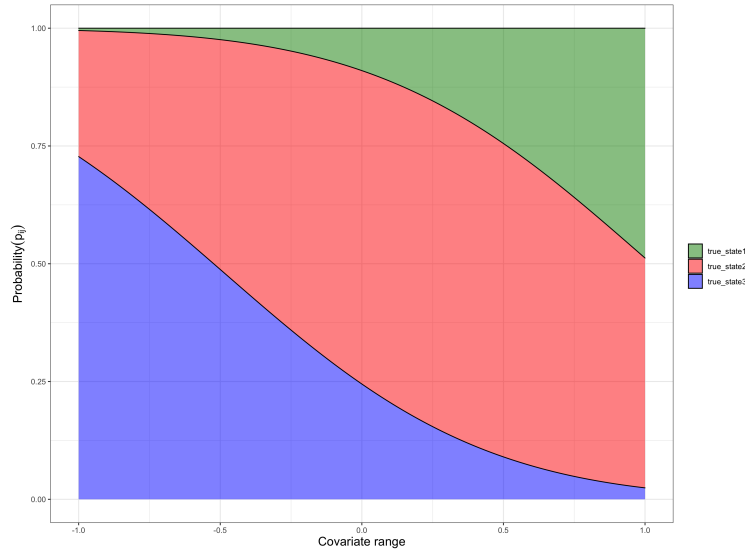


Figure 1: Probability of verified (true) states from the simulation study parameter choices over the given covariate domain.

The confusion matrix used for the simulation study is presented in Table 2. The values were chosen such that it is impossible to misclassify state 3 as category 1 or 2.

Verified states	Observed category			
	1	2	3	4
1	0.7	0.05	0.13	0.12
2	0.10	0.80	0.05	0.05
3	0	0	0.9	0.10

Table 2: The classification probabilities (confusion matrix) used in the simulation study.

The simulation was repeated 1000 times, and for each run the true intensity was simulated using the same parameters as defined above. This was done to ensure that the results analysed would not be due to randomness.

### 2.4.1 Fitting the model and estimating the parameters

The model was fitted with MCMC using the Bayesian modelling framework. All the analyses were done using *R* software [Chambers, 2008]. The mcmc was run with the nimble package [de Valpine et al., 2017]. The priors for  $\beta_{0i}$  and  $\beta_{1i}$  for  $i = 1, 2, \dots, S$  states were chosen from a normal distribution with mean 0 and standard deviation of 1, i.e.  $\beta_{0i} \sim N(0, 1)$  and  $\beta_{1i} \sim N(0, 1)$ . The priors of the confusion matrix ( $\Omega$ ) were chosen from the Dirichlet distribution with parameter alpha equal to 1.

The chains were run for  $5 \times 10^5$  number of iterations for 3 chains, and the first  $2 \times 10^5$  iterations were chosen as the burn-in. The fitted model was ignored when the model did not converge (looking at trace plots and with a Gelman statistic value greater than 1.04 Brooks and Gelman [1998]), showed an error in the running of the model as well as having a small effective sample size. The convergence was done with the 'ggmcmc' package [Fernández-i Marín, 2016] in *R*.

To check how well we estimated the parameters in the model, the model parameters  $\beta_0$ ,  $\beta_1$  and  $\Omega$  were evaluated by estimating the root mean squared error (RMSE) and coverage of the estimated parameters. Better results would be parameters with lower RMSE and higher coverage. The results were presented with the 'ggplot' package [Wickham et al., 2016] in *R*.

## 2.5 Application to data

We demonstrate application of the model using a butterfly dataset that was downloaded from iNaturalist. iNaturalist is a Citizen Science platform that hosts over ninety million occurrence observations with well over a million observers (Website <https://www.inaturalist.org/> visited on 18th March 2022). Observers collect occurrence records and upload their observations with images, sounds and other valuable information that allow for verification [Matheson, 2014]. The species name provided by the observer is known as the species guess or reported species. The observations reported goes through a community verification, and is accepted when two-thirds of users verifying the observation agree upon the taxon identification [Ueda, 2020]. Community verified observations are termed "research grade", and we will in the following refer to these observations as the true species. Observations without community verification are termed as "needs id".

The butterfly data used for the demonstration was collected from United States of America in the period from 2019 to 2020. Specifically we selected the butterfly species: the viceroy (*Limenitis archippus*), the queen butterfly (*Danaus gilippus*) and the monarch butterfly (*Danaus plexippus*). Any other species apart from *Limenitis archippus* in the genus

*Limenitis* was re-coded as "other Limenitis", other species in the *Danaus* genus was also re-coded as "other Dananus" and any other species outside these species categories were denoted as "other". The distribution of the species to be analysed are presented in Figure 2.

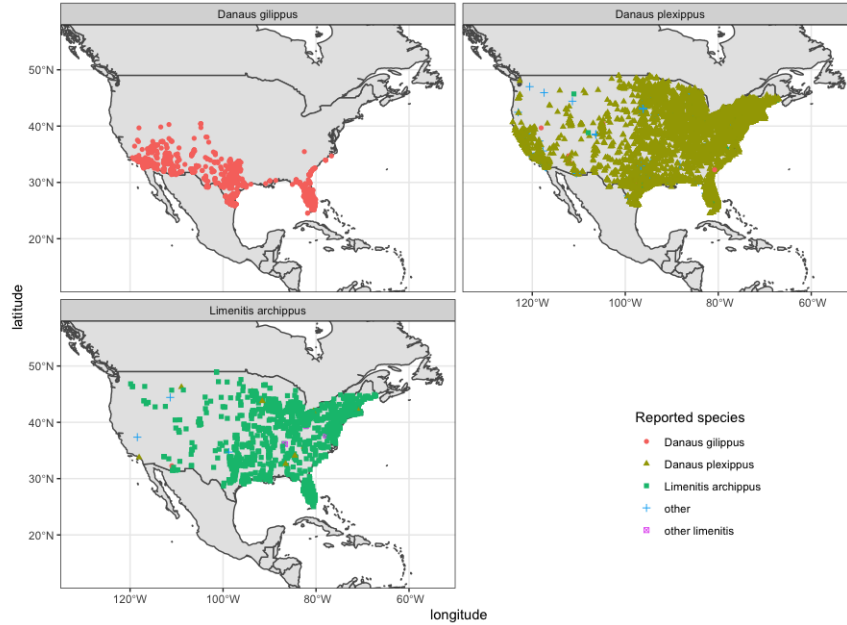


Figure 2: The subset of the butterfly dataset downloaded from iNaturalist used for the analysis. The data are faceted according to the true species name and coloured according to the reported category.

Climatic factors such as precipitation, temperature and elevation of the area are major drivers behind the population dynamics of the North American Monarch butterfly [Brym et al., 2020]. The butterfly has an annual multigenerational migration cycle and for every fall, those at the Rocky Mountains travel great distances to the summer breeding grounds, where most of them aggregate [Culbertson et al., 2022]. Culbertson et al. [2022] noted that climatic changes affects this migration as Zipkin et al. [2012] adds that this is a major driver in its population dynamics. The queen butterfly and viceroy butterfly are co-mimics and have great resemblance because of their patterns [Prudic et al., 2019, Ritland, 1991]. This suggests that the abundance of one of them affects the other in the location.

We developed a spatial model where we used elevation from the Bioclim data in the raster package [Hijmans et al., 2013], month, number of agreements and number of disagreements as covariates. The true intensity of the verified species were estimated using INLA [Rue et al., 2019] and the classification probabilities in nimble [de Valpine et al., 2017]. The details can be found in the code in the Supplementary information.

The true species for the data that needed verification in 2019 was predicted by retrieving the species with the highest posterior predicted probability. Butterfly data for the states of Florida, Montana, Minnesota, Wisconsin, Illinois and Indiana in 2020 were selected and the true species for both the data that needed verification and those with community verification were predicted by retrieving the species with the highest posterior predicted probability.



### 3 Results

#### 3.1 Simulation study

##### 3.1.1 Intensity of verified states

Table 3 presents the estimates of the parameters from the 1000 simulations summarised by using the median, root mean square error (RMSE) and the coverage. The contrast in the estimates, using state 1 as the reference state, were less biased with a small root mean square error and high coverage as can be observed from Table 3. This is what we expected considering that the multinomial logit regression was used for the verified states.

Parameters	True value	Estimate	RMSE	Coverage
$\beta_{02} - \beta_{01}$	2	1.89	0.046	0.94
$\beta_{03} - \beta_{01}$	1	0.91	0.058	0.96
$\beta_{12} - \beta_{11}$	-2	-1.76	0.058	0.97
$\beta_{13} - \beta_{11}$	-4	-3.67	0.1471	0.91
$\Omega_{11}$	0.7	0.69	$3.8 \times 10^{-3}$	0.94
$\Omega_{21}$	0.10	0.10	$3.63 \times 10^{-4}$	0.95
$\Omega_{31}$	0.00	0.00	$1.6 \times 10^{-9}$	—
$\Omega_{12}$	0.05	0.05	$8.26 \times 10^{-4}$	0.92
$\Omega_{22}$	0.80	0.79	$6.66 \times 10^{-4}$	0.96
$\Omega_{32}$	0.00	0.00	$1.60 \times 10^{-9}$	-
$\Omega_{13}$	0.13	0.13	$1.84 \times 10^{-3}$	0.96
$\Omega_{23}$	0.05	0.05	$2.21 \times 10^{-4}$	0.95
$\Omega_{33}$	0.90	0.9	$7.28 \times 10^{-4}$	0.97
$\Omega_{14}$	0.12	0.12	$1.99 \times 10^{-3}$	0.93
$\Omega_{24}$	0.05	0.05	$2.09 \times 10^{-4}$	0.96
$\Omega_{34}$	0.10	0.10	$7.13 \times 10^{-4}$	0.97

Table 3: True values and posterior summaries of estimates of parameters from the simulation study. There are no coverage values for  $\Omega_{31}$  and  $\Omega_{32}$  because the true value can never be within the confidence interval. All the results are converted to two decimal places.

##### 3.1.2 Classification probabilities

Table 3 provides estimates of the parameter of interest, the classification probabilities ( $\Omega$ ). The estimates of classification probabilities were almost unbiased with a small root mean square error and higher coverage for most cases. Even though we obtained relative abundances for the verified states as discussed above, the model proposed provides precise and accurate estimates of the classification probabilities. This suggests that the estimation of the classification probabilities is somewhat independent of the model for the intensity or verified species. So one can estimate the classification probability precisely and accurately for any given distribution model the verified species takes.

### 3.1.3 Predicting the verified states

As explained in Section 2 above, the true verified states can be predicted from the reported states and the classification probabilities. The posterior probability of verified states  $s$  given the reported category is shown in Figure 3, for  $i = 1, 2, 3$  and  $K = 4$  categories.

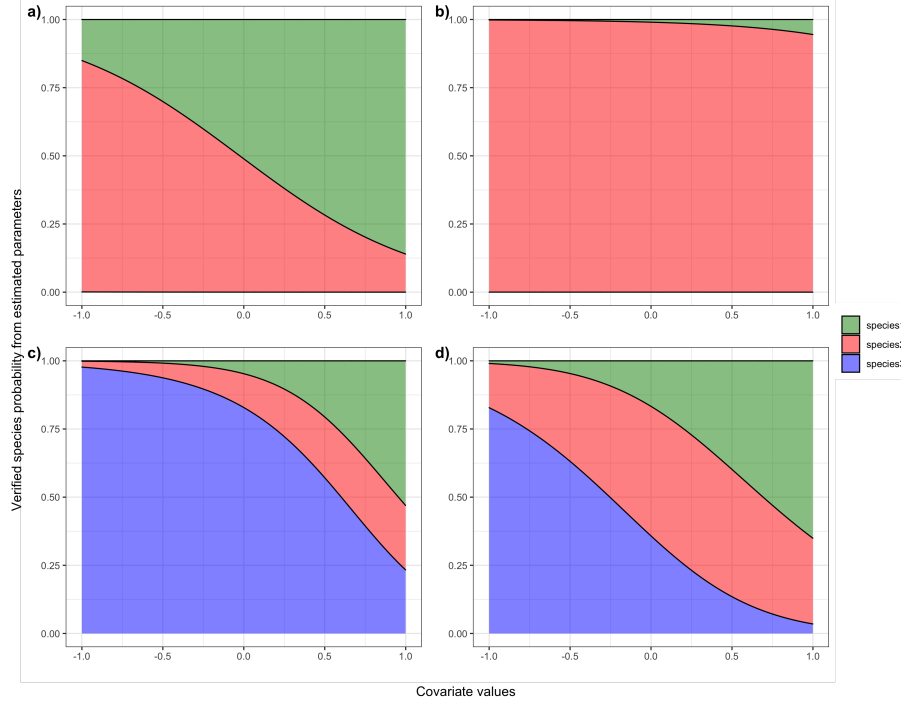


Figure 3: The figure shows the predicted posterior probability of observing verified states given that reported category is a) category 1, b) category 2, c) category 3 and d) category 4. The green shaded colour shows the region of coverage for the posterior probability of species 1, red colour for species 2 and blue colour for species 3.

The results in Figure 3 showed that the posterior probability of identifying any of the verified states depends on the location of the state as well as its misclassification probabilities. For example, assuming that category 1 is reported, the posterior probability curve of the state 1 was shifted down since it has a higher probability of being the true state over the entire covariate range and suggesting an impossibility for the reported category to be verified state 3 (from Figure 3 a)), even for regions where true state 3 was higher in the true simulation in Figure 1. Similar deductions can be made when categories 2, 3 and 4 were reported.

## 3.2 Data Application

The results for the classification probabilities and the parameters of the model are presented in the Supplementary information. Although the viceroy and the queen are co-mimics [Ritland, 1991], the misclassification for reporting viceroy as queen butterfly and the converse was negligible, suggesting that the citizen scientists were able to distinguish between them.

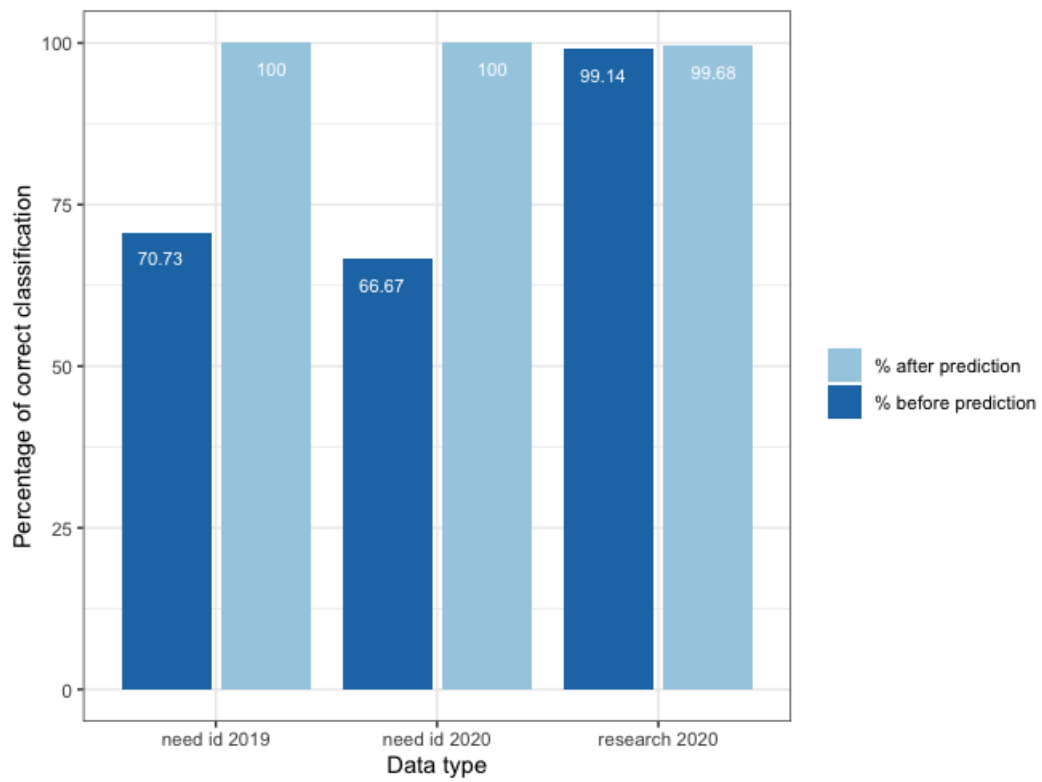


Figure 4: Percentage of correct classification for the data that needed verification in 2019 ("need id 2019") and 2020 ("need id 2020") and those with research grade in 2020 ("research 2020") before (deep blue colour) and after (light blue colour) the posterior prediction of the true species. The percentages for each category are presented in their corresponding bars.

The proportion of correct classification for the true species identity that needed verification in 2019 increased from 70.73% before prediction to 100% after the posterior prediction with the model, as can be seen from Figure 4. The proportion of correct classification for the species that needed verification in 2020 increased from 66.67% before prediction to 100% after the prediction (Figure 4). For the data with verified true species identity, there was a 0.5% increase from prediction using our model(Figure 4).

## 4 Discussion

The current study has provided a general framework to account for misclassification in the multi-species distribution models. The model can be used to predict the true species, and how this prediction depends on the misclassification probabilities and the actual distribution.

The simulation study demonstrated that if we think we have observed a species that is difficult to identify, what it actually is will depend on where we observe it. This matches the behaviour of ornithologists when asked to identify a bird from a photo: their first question is usually "where was the photograph taken?" (GrrlScientist, *pers. comm*)

With the model for the true intensity (equation (1)) being a multinomial logit model, the model returned the contrasts of the estimates of the parameters relative to a reference species. Thus, any model that is assumed for the generation of the verified states, such as point processes, etc as well as any model for estimating the ecological parameters such as joint species distribution models should be feasible for this framework.

Following from the argument in the previous paragraph, the model framework proposed here can be used to identify probable misidentifications, by flagging observations that are improbable of being correctly classified. Our simulation study showed that in real word situations, if the species are relatively likely to occur in a region, the posterior probability favours the species with the highest classification in that reported category, and the vice versa. Thus we can not only flag that a species may be misidentified (if the identification has a low probability of being correct), but we can also suggest the correct identification. With the large number of observations coming into data portals such as iNaturalist, a model such as ours can form the basis of automated screening of observations, to identify those which might need manual checking.

Our simulation study stressed on the importance of accounting for misclassification in multi-species models just as was noted in previous literature [Kéry and Royle, 2020, Chambert et al., 2015]. The posterior predicted probability for the reported states heavily relies on the weights from the misclassification probability. Failure to account for that would mean our posterior probability would be the same as the probability of the verified state. This would lead to the underestimation of the prediction of the ranges of the coverage and possibly abundance in the verified states [Molinari-Jobin et al., 2012].

The model framework proposed does not just account for misidentifications, but identifications at different tax-

onomic levels. This implies that CS data analysts can make use of more data, especially when the data have some missing identifications for a taxonomic level, which will sometimes be more informative. We have shown that our model can improve an identification based on the location, for example, an unidentified butterfly from our data is unlikely to be a queen butterfly if it is observed in Wisconsin, because queen butterfly is not found in that state.

Our model showed a 29% increase in the number of correct identifications for the data that needed verification in 2019, 33% increase in the number of correct identifications for the data that needed verification in 2020 and a non-significant increase in 2020 for data with community verification. Although the queen and viceroy butterflies have a greater resemblance between their ventral patterns and co-mimic each other [Ritland, 1991], citizen scientists appeared to be able to correctly classify these two species for all the months. Although the model proposed was able to predict the true species given the reported category to a higher degree, we suggest that the community verification is a necessity to get the true species information.

The model framework is flexible and can be generalised into any joint species distribution models. Although we use a point process model for the true distribution, our misclassification model can be plugged into similar model that estimates either abundance or incidence. And because our model does not assume the true states are the only possible reported states, it allows for a lot of flexibility in what is reported (for example, it could be developed to take DNA sequences and match them to species). This will make it possible to analyse Citizen Science data alongside other data sources to account for misclassification at any level in the data collection process.

## References

- Tomas J Bird, Amanda E Bates, Jonathan S Lefcheck, Nicole A Hill, Russell J Thomson, Graham J Edgar, Rick D Stuart-Smith, Simon Wotherspoon, Martin Krkosek, Jemina F Stuart-Smith, et al. Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, 173:144–154, 2014.
- John DJ Clare, Philip A Townsend, and Benjamin Zuckerberg. Generalized model-based solutions to false-positive error in species detection/nondetection data. *Ecology*, 102(2):e03241, 2021.
- Marc Kéry and J Andrew Royle. *Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and Advanced Models*. Academic Press, 2020.
- Wilson J Wright, Kathryn M Irvine, Emily S Almberg, and Andrea R Litt. Modelling misclassification in multi-species acoustic data when estimating occupancy and relative activity. *Methods in Ecology and Evolution*, 11(1): 71–81, 2020.
- Thierry Chambert, David AW Miller, and James D Nichols. Modeling false positive detections in species occurrence data under different study designs. *Ecology*, 96(2):332–339, 2015.
- J Andrew Royle and William A Link. Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, 87(4):835–841, 2006.
- Thierry Chambert, Evan H Campbell Grant, David AW Miller, James D Nichols, Kevin P Mulder, and Adrienne B Brand. Two-species occupancy modelling accounting for species misidentification and non-detection. *Methods in Ecology and Evolution*, 9(6):1468–1477, 2018a.
- Anja Molinari-Jobin, M Kéry, E Marboutin, P Molinari, I Koren, C Fuxjäger, Ch Breitenmoser-Würsten, S Wölfl, M Fasel, I Kos, et al. Monitoring in the presence of species misidentification: the case of the eurasian lynx in the alps. *Animal Conservation*, 15(3):266–273, 2012.
- Norwegian Biodiversity Information Centre. Artsorakel (version 2.01.04) [mobile app] google play., 2022. URL <https://play.google.com/store/apps/details?id=no.artsdatabanken.orakel>.
- Christie A Matheson. inaturalist. *Reference Reviews*, 2014.
- Marco Willi, Ross T Pitman, Anabelle W Cardoso, Christina Locke, Alexandra Swanson, Amy Boyer, Marten Veldthuis, and Lucy Fortson. Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, 10(1):80–91, 2019.
- Anisha Keshavan, Jason D Yeatman, and Ariel Rokem. Combining citizen science and deep learning to amplify expertise in neuroimaging. *Frontiers in neuroinformatics*, 13:29, 2019.

- Zakaria Saoud, Colin Fontaine, Grégoire Lois, Romain Julliard, and Iandry Rakotoniaina. Miss-identification detection in citizen science platform for biodiversity monitoring using machine learning. *Ecological Informatics*, 60:101176, 2020.
- Noel Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.
- Ian W Renner, Jane Elith, Adrian Baddeley, William Fithian, Trevor Hastie, Steven J Phillips, Gordana Popovic, and David I Warton. Point process models for presence-only analysis. *Methods in Ecology and Evolution*, 6(4):366–379, 2015.
- John R Sauer, William A Link, Jane E Fallon, Keith L Pardieck, and David J Ziolkowski. The north american breeding bird survey 1966–2011: summary analysis and species accounts. *North American Fauna*, (79 (79)):1–32, 2013.
- GD Steel et al. Relation between poisson and multinomial distributions. 1953.
- Thierry Chambert, J Hardin Waddle, David AW Miller, Susan C Walls, and James D Nichols. A new framework for analysing automated acoustic species detection data: Occupancy estimation and optimization of recordings post-processing. *Methods in Ecology and Evolution*, 9(3):560–570, 2018b.
- John Chambers. *Software for data analysis: programming with R*. Springer Science & Business Media, 2008.
- Perry de Valpine, Daniel Turek, Christopher J Paciorek, Clifford Anderson-Bergman, Duncan Temple Lang, and Rastislav Bodik. Programming with models: writing statistical algorithms for general model structures with nimble. *Journal of Computational and Graphical Statistics*, 26(2):403–413, 2017.
- Stephen P Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.
- Xavier Fernández-i Marín. ggmcmc: Analysis of mcmc samples and bayesian inference. *Journal of Statistical Software*, 70:1–20, 2016.
- Hadley Wickham, Winston Chang, and Maintainer Hadley Wickham. Package ‘ggplot2’. *Create Elegant Data Visualisations Using the Grammar of Graphics. Version*, 2(1):1–189, 2016.
- K Ueda. inaturalist research-grade observations. inaturalist. org. *Occurrence dataset <https://doi.org/10.15468/ab3s5x> [accessed via GBIF.org on 26 May 2020, <https://www.gbif.org/occurrence/2294559370>]*, 2020.
- Matthew Z Brym, Cassandra Henry, Shannon P Lukashow-Moore, Brett J Henry, Natasja Van Gestel, and Ronald J Kendall. Prevalence of monarch (*danaus plexippus*) and queen (*danaus gilippus*) butterflies in west texas during the fall of 2018. *BMC ecology*, 20(1):1–13, 2020.

- Katherine A Culbertson, Mark S Garland, Richard K Walton, Louise Zemaitis, and Victoria M Pocius. Long-term monitoring indicates shifting fall migration timing in monarch butterflies (*danaus plexippus*). *Global Change Biology*, 28(3):727–738, 2022.
- Elise F Zipkin, Leslie Ries, Rick Reeves, James Regetz, and Karen S Oberhauser. Tracking climate impacts on the migratory monarch butterfly. *Global Change Biology*, 18(10):3039–3049, 2012.
- Kathleen L Prudic, Barbara N Timmermann, Daniel R Papaj, David B Ritland, and Jeffrey C Oliver. Mimicry in viceroy butterflies is dependent on abundance of the model queen butterfly. *Communications biology*, 2(1):1–9, 2019.
- David B Ritland. Revising a classic butterfly mimicry scenario: demonstration of müllerian mimicry between florida viceroys (*limenitis archippus floridensis*) and queens (*danaus gilippus berenice*). *Evolution*, 45(4):918–934, 1991.
- Robert J Hijmans, J van Etten, M Mattiuzzi, M Sumner, JA Greenberg, OP Lamigueiro, A Bevan, EB Racine, and A Shortridge. Raster package in r, 2013.
- H Rue, F Lindgren, D Simpson, S Martino, E Teixeira Krainski, H Bakka, A Riebler, and GA Fuglstad. Inla: Full bayesian analysis of latent gaussian models using integrated nested laplace approximations. *R package version*, 19(03), 2019.